



# Full wwPDB X-ray Structure Validation Report i

Mar 31, 2014 – 07:53 PM EDT

PDB ID : 4JLG  
Title : SETD7 in complex with inhibitor (R)-PFI-2 and S-adenosyl-methionine  
Authors : Dong, A.; Wu, H.; Zeng, H.; El Bakkouri, M.; Barsyte, D.; Vedadi, M.; Tatlock, J.; Owen, D.; Bunnage, M.; Bountra, C.; Arrowsmith, C.H.; Edwards, A.M.; Brown, P.J.; Structural Genomics Consortium (SGC)  
Deposited on : 2013-03-12  
Resolution : 1.90 Å(reported)

This is a full wwPDB validation report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

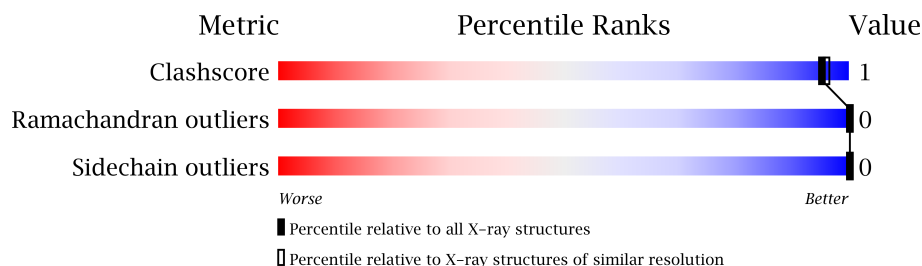
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1439
EDS	:	<b>FAILED</b>
Percentile statistics	:	21963
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22978

# 1 Overall quality at a glance

The reported resolution of this entry is 1.90 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	79885	4465 (1.90-1.90)
Ramachandran outliers	78287	4413 (1.90-1.90)
Sidechain outliers	78261	4414 (1.90-1.90)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	264	
1	B	264	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 4374 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

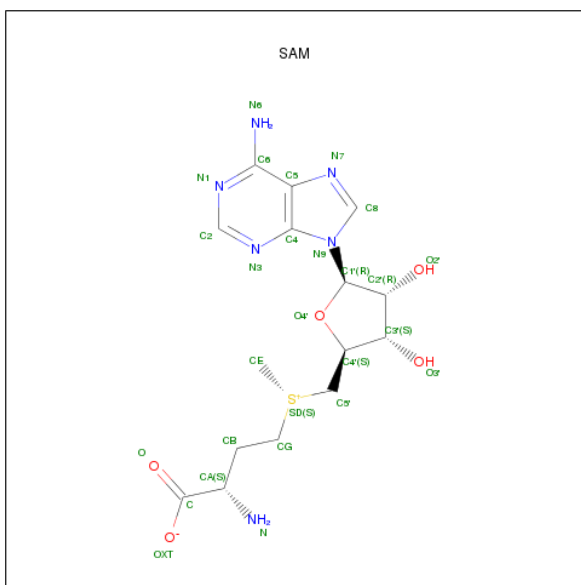
- Molecule 1 is a protein called Histone-lysine N-methyltransferase SETD7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	247	Total	C	N	O	S	0	4	0
			1904	1211	307	374	12			
1	B	246	Total	C	N	O	S	0	7	0
			1921	1225	311	373	12			

There are 12 discrepancies between the modelled and reference sequences:

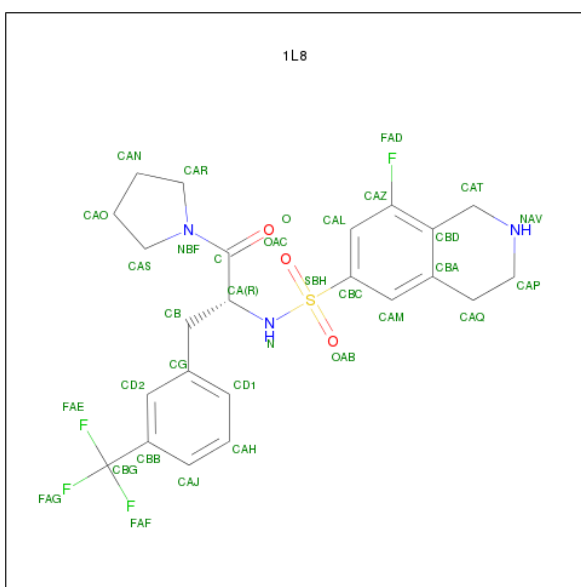
Chain	Residue	Modelled	Actual	Comment	Reference
A	367	HIS	-	EXPRESSION TAG	UNP Q8WTS6
A	368	HIS	-	EXPRESSION TAG	UNP Q8WTS6
A	369	HIS	-	EXPRESSION TAG	UNP Q8WTS6
A	370	HIS	-	EXPRESSION TAG	UNP Q8WTS6
A	371	HIS	-	EXPRESSION TAG	UNP Q8WTS6
A	372	HIS	-	EXPRESSION TAG	UNP Q8WTS6
B	367	HIS	-	EXPRESSION TAG	UNP Q8WTS6
B	368	HIS	-	EXPRESSION TAG	UNP Q8WTS6
B	369	HIS	-	EXPRESSION TAG	UNP Q8WTS6
B	370	HIS	-	EXPRESSION TAG	UNP Q8WTS6
B	371	HIS	-	EXPRESSION TAG	UNP Q8WTS6
B	372	HIS	-	EXPRESSION TAG	UNP Q8WTS6

- Molecule 2 is S-ADENOSYLMETHIONINE (three-letter code: SAM) (formula: C<sub>15</sub>H<sub>22</sub>N<sub>6</sub>O<sub>5</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total 27	C 15	N 6	O 5	S 1	0	0
2	B	1	Total 27	C 15	N 6	O 5	S 1	0	0

- Molecule 3 is 8-FLUORO-N-{(2R)-1-OXO-1-(PYRROLIDIN-1-YL)-3-[3-(TRIFLUOROMETHYL)PHENYL]PROPAN-2-YL}-1,2,3,4-TETRAHYDROISOQUINOLINE-6-SULFONAMIDE (three-letter code: 1L8) (formula:  $C_{23}H_{25}F_4N_3O_3S$ ).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	C	F	N	O	S	0	0
			34	23	4	3	3	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	B	1	Total	C	F	N	O	S	0	0
			34	23	4	3	3	1		

- Molecule 4 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	8	Total	X	0	0
			8	8		
4	A	17	Total	X	0	0
			17	17		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	200	Total	O	0	5
			205	205		
5	B	194	Total	O	0	3
			197	197		

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

Note EDS failed to run properly.

- Molecule 1: Histone-lysine N-methyltransferase SETD7

Chain A: 



- Molecule 1: Histone-lysine N-methyltransferase SETD7

Chain B: 



## 4 Data and refinement statistics i

EDS failed to run properly - this section will therefore be incomplete.

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	46.74Å 57.98Å 64.93Å 83.33° 71.42° 70.93°	Depositor
Resolution (Å)	30.75 – 1.90	Depositor
% Data completeness (in resolution range)	94.8 (30.75-1.90)	Depositor
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.53 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.211 , 0.237	Depositor
Wilson B-factor (Å <sup>2</sup> )	23.9	Xtriage
Anisotropy	0.223	Xtriage
Estimated twinning fraction	0.000 for -h,-k,-h+l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 45881 reflections	Xtriage
Total number of atoms	4374	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.80% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: UNX, SAM, 1L8

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.44	0/1964	0.52	0/2676
1	B	0.43	0/1992	0.52	0/2714
All	All	0.43	0/3956	0.52	0/5390

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1904	0	1786	6	0
1	B	1921	0	1814	4	0
2	A	27	0	22	0	0
2	B	27	0	22	0	0
3	A	34	0	25	1	0
3	B	34	0	25	0	0
4	A	17	0	0	0	0
4	B	8	0	0	0	0
5	A	205	0	0	0	0
5	B	197	0	0	1	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	4374	0	3694	10	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 1.

All (10) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:117:GLY:O	1:A:132:GLU:HG2	2.08	0.54
1:A:266:THR:O	3:A:402:1L8:H8	2.10	0.52
1:B:168:LYS:HD2	5:B:512:HOH:O	2.14	0.48
1:A:305:TYR:CZ	1:A:333[B]:VAL:HG21	2.49	0.47
1:A:244:PHE:O	1:A:290:SER:HB2	2.17	0.44
1:A:129:LEU:HD21	1:A:162:GLY:HA2	1.99	0.44
1:B:344[A]:LYS:HB2	1:B:344[A]:LYS:NZ	2.34	0.43
1:A:310:HIS:CE1	1:A:312:ARG:HB3	2.55	0.42
1:B:194:ASP:OD2	1:B:203[A]:THR:HG23	2.20	0.42
1:B:297:HIS:HA	1:B:333:VAL:O	2.21	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	249/264 (94%)	241 (97%)	8 (3%)	0	100	100
1	B	251/264 (95%)	242 (96%)	9 (4%)	0	100	100
All	All	500/528 (95%)	483 (97%)	17 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	203/227 (89%)	203 (100%)	0	100	100
1	B	206/227 (91%)	206 (100%)	0	100	100
All	All	409/454 (90%)	409 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	192	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 29 ligands modelled in this entry, 25 are unknown - leaving 4 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected

value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	SAM	A	401	-	26,29,29	0.96	2 (7%)	38,42,42	2.08	6 (15%)
3	1L8	A	402	-	37,37,37	2.93	8 (21%)	55,55,55	2.02	6 (10%)
2	SAM	B	401	-	26,29,29	0.91	2 (7%)	38,42,42	1.99	5 (13%)
3	1L8	B	402	-	37,37,37	2.91	7 (18%)	55,55,55	2.10	5 (9%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	SAM	A	401	-	-	0/13/33/33	0/1/3/3
3	1L8	A	402	-	-	0/29/43/43	0/2/4/4
2	SAM	B	401	-	-	0/13/33/33	0/1/3/3
3	1L8	B	402	-	-	0/29/43/43	0/2/4/4

All (19) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	402	1L8	OAB-SBH	9.76	1.54	1.43
3	B	402	1L8	OAB-SBH	9.69	1.54	1.43
3	B	402	1L8	OAC-SBH	9.29	1.53	1.43
3	A	402	1L8	OAC-SBH	8.92	1.53	1.43
3	A	402	1L8	CAT-CBD	-7.69	1.41	1.51
3	B	402	1L8	CAT-CBD	-7.65	1.41	1.51
3	A	402	1L8	CAQ-CBA	-5.40	1.41	1.51
3	B	402	1L8	CAQ-CBA	-5.07	1.41	1.51
3	A	402	1L8	CB-CG	-4.09	1.41	1.51
3	B	402	1L8	CB-CG	-3.80	1.41	1.51
3	B	402	1L8	CBG-CBB	-3.11	1.42	1.49
2	B	401	SAM	C2-N3	3.02	1.38	1.32
2	A	401	SAM	C2-N3	2.96	1.38	1.32
3	A	402	1L8	CBG-CBB	-2.68	1.43	1.49
3	A	402	1L8	FAD-CAZ	-2.56	1.29	1.35
3	B	402	1L8	FAD-CAZ	2.49	1.41	1.35
2	A	401	SAM	C2-N1	2.23	1.38	1.33
2	B	401	SAM	C2-N1	2.16	1.38	1.33
3	A	402	1L8	FAG-CBG	-2.01	1.24	1.32

All (22) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	402	1L8	CAL-CAZ-CBD	-10.99	120.31	124.74
2	A	401	SAM	N3-C2-N1	-10.63	119.82	128.71
2	B	401	SAM	N3-C2-N1	-9.68	120.61	128.71
3	A	402	1L8	CAL-CAZ-CBD	-9.58	120.88	124.74
3	A	402	1L8	OAC-SBH-OAB	-7.99	108.80	119.55
3	B	402	1L8	OAC-SBH-OAB	-7.45	109.53	119.55
3	A	402	1L8	CA-N-SBH	-3.95	112.88	121.34
2	B	401	SAM	N3-C4-N9	3.69	132.10	125.43
3	A	402	1L8	CBC-SBH-N	3.38	112.24	107.76
2	A	401	SAM	N3-C4-N9	3.33	131.45	125.43
3	B	402	1L8	CBC-SBH-N	3.27	112.09	107.76
3	B	402	1L8	CA-N-SBH	-2.93	115.08	121.34
2	A	401	SAM	C4-C5-N7	-2.66	107.24	109.52
3	B	402	1L8	FAD-CAZ-CBD	2.66	121.25	117.52
3	A	402	1L8	OAB-SBH-N	2.50	111.41	106.84
2	B	401	SAM	C4-C5-N7	-2.48	107.39	109.52
2	B	401	SAM	C5-C4-N3	-2.47	120.33	125.70
3	A	402	1L8	FAD-CAZ-CBD	2.38	120.87	117.52
2	A	401	SAM	C5-C4-N3	-2.17	120.97	125.70
2	B	401	SAM	C2-N3-C4	2.14	120.11	114.01
2	A	401	SAM	C2-N3-C4	2.12	120.05	114.01
2	A	401	SAM	N7-C8-N9	-2.03	108.61	114.36

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

EDS failed to run properly - this section will therefore be empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

EDS failed to run properly - this section will therefore be empty.

### 6.3 Carbohydrates ⓘ

EDS failed to run properly - this section will therefore be empty.

### 6.4 Ligands ⓘ

EDS failed to run properly - this section will therefore be empty.

### 6.5 Other polymers ⓘ

EDS failed to run properly - this section will therefore be empty.